Home Advantage: Exploring Evidence for Explanations Using Cricket Data

by J. James Reade
Home Advantage: Exploring evidence for explanations using cricket data *

J. James Reade
February 5, 2018

Abstract

Home advantage is the observed regularity that participants in contests win more often than their relative quality would suggest when playing at home. This paper reviews the literature and illustrates a number of the potential explanations using a huge dataset of cricket matches. Explanations for the home advantage can be summarised into four headings: crowd, familiarity, travel and rules. Evidence increasingly points towards the role that officials play, yet other explanations cannot necessarily be ruled out.

This paper has been prepared as a chapter for the Handbook of Sports Economics, to be published by SAGE, edited by Downward, P., Frick, B., Humphreys, B.R, Pawlowski, T., Ruseski, J. and Soebbing, B.

JEL Classification: Z2, C20, L83.

Keywords: Home advantage, contests, sport.

1 Introduction

Economics, the study of outputs for a set of decisions regarding scarce inputs, is thus interested in systematic patterns in outputs. Sport provides a range of outputs, or outcomes, that are very measurable, and attract great interest.

Rottenberg (1956) was the first to note the joint production involved in sport, as both teams contribute to the output, a contest, that is demanded by consumers. Additionally, officials contribute. Officials are agents responsible for regulating activity to ensure that participation is fair. Such officials are granted power to penalise contestants who do not act fairly. In football, for example, a team of referees operate on the field during a contest, while Football Associations operate more broadly, stipulating rules and regulations to be enforced.

One incredibly persistent pattern in sporting outcomes is that the home team wins more often than it ought to given its strength relative to the visiting team. If two identical teams from different venues played against each other, the team playing at their venue would win more than 50% of the time. Pollard and Pollard (2005b) identify its existence over more than a century.

In this paper we explore the various competing mechanisms proposed for its existence, and pay particular attention to explanations that involve officials. We illustrate the results found in the literature

*Thanks to the members of the Association of Cricket Statisticians and Historians, Sarah Jewell, Liam Lenten and Owen Gittings for their invaluable insight and comments during the writing of this paper. All remaining errors are mine.
by analysing a dataset hitherto not before used in the analysis of home advantage, namely every single cricket match recorded on the Cricket Archive website between 1731 and 2017.

In Section 2 we introduce the concept of home advantage in more detail, and outline a method of identifying it using data, in Section 3 we review the key papers on the subject of home advantage, using cricket data to illustrate a number of the points. Section 4 concludes.

2 The Concept of Home Advantage

We define home advantage as the home team winning more often than it ought to given the relative ability of the teams competing. Equally, however, for sport to generate interest, it is argued there must be sufficient uncertainty of outcome. The uncertainty of outcome hypothesis, originally formulated by Rottenberg [1956], implies that outcomes should not be exactly proportional to the abilities of participants, but that there ought to be some degree of variation around expected outcomes given abilities of participants.

Provided that the ability of participants can be measured, we can identify home advantage. If we define the outcome of event $i$ to be $y_i$, and the relative strength of the home team to be a function of the strengths of the two teams competing, $x_i = f(x_{H,i}, x_{A,i})$, then we could run a simple regression model of:

$$y_i = \alpha + \beta x_i + e_i.$$  \hspace{1cm} (1)

Here, $\beta$ would represent the rate at which outcomes depend on relative ability. The error term $e_i$ represents the extent to which there is random variation around relative abilities and outcomes. Given the existence of $\alpha$, $e_i$ must be mean zero, and hence if it had variance zero, this would imply no uncertainty in outcomes. The larger is the residual variance, the greater is the variation around relative abilities: greater uncertainty of outcome.

This approach equates to a Mincer and Zarnowitz [1969] regression test on the forecasts produced by the relative strengths of the two teams competing. The Mincer-Zarnowitz approach is commonly used to evaluate the efficiency of forecasts: whether they incorporate all information available at the forecast origin.

As with the Mincer-Zarnowitz test, we are interested in the values of the $\alpha$ and $\beta$ coefficients, and the absence of home advantage implies that $\alpha = 0$ and $\beta = 1$, or that the relative ability of teams is the efficient predictor of match outcomes (given uncertainty of outcomes). In that case, $E(y_i) = x_i$, the expected outcome is equal to the relative quality of the two teams. The $\alpha$ coefficient is the increase in the probability of a home victory conditional on the relative strengths remaining constant. In the absence of home advantage, $\alpha = 0$. The $\beta$ slope coefficient regulates the change in the probability of a home win for a change in the relative strengths of the teams competing.

If $\beta < 1$ then this implies that as the home team becomes relatively stronger, their likelihood of winning increases at a less than proportional rate. If $\beta > 1$ then as the home team becomes relatively stronger, their likelihood of winning increases at a more than proportional rate. With $\beta < 1$ this implies that home advantage favours weaker teams, whereas $\beta > 1$ implies that home advantage favours stronger teams, although in both cases this is contingent on the $\alpha$ coefficient.

---

1We would require that $\partial x_i/\partial x_{H,i} > 0$ and $\partial x_i/\partial x_{A,i} < 0$, and further that the function is symmetric such that $f(x_{H,i}, x_{A,i}) = f(x_{A,i}, x_{H,i})$. We require only that $x_i$ is ordinal and not cardinal; it needs to be able to order any team pairing in terms of weaker and stronger.
The $\beta < 1$ case could be argued to be consistent with profit maximising sports leagues, since it implies weaker home sides win more often than their relative ability suggests, thus increasing interest within the local market for attendance. Additionally, if strong home teams always beat weaker sides, this might lessen interest in such matches, even from the fans of the stronger team. Note that $\beta < 1$ is necessary but not sufficient for home advantage; on its own, this simply implies stronger teams win less often than they should. In addition, an $\alpha$ coefficient such that the point where the regression line intersects the 45-degree line above 0.5 is required for home advantage to exist, since the Elo prediction for two teams of equal quality is 0.5.

The nature of the $y_i$ variable has not been determined, and this has implications for the most appropriate resulting regression model. A continuous variable is difficult to imagine given the discrete nature of the outcomes of many sporting events (home team wins or loses). For some events, outcomes are expressed in non-comparable units (runs or wickets).\(^2\) Even for outcomes that are more continuous, the margins of victory that would determine a maxima or minima are difficult, if not impossible to determine.\(^3\)

Conventionally we may think about $y_i$ as being zero for an away team victory, one for a home team win, and some third value (say, a half) for a draw:

$$y_i = \begin{cases} 
1 & \text{if home team wins,} \\
0.5 & \text{if match drawn,} \\
0 & \text{if away team wins,}
\end{cases} \quad (2)$$

We adopt (2) as our dependent variable when measuring home advantage in this paper.\(^4\) Koop (2004) consider multinomial and ordered models for baseball, Vlastakis et al. (2009), Goddard and Asimakopoulos (2004), and Goddard (2005) for football, and Scarf and Shi (2005) and Akhtar and Scarf (2012) for cricket. In general, it might be argued that the distinction between a home or away win has no obvious ordering, and hence that a multinomial model is appropriate. However, when the metric is the relative performance of the home team, then the ordering is clear: away win, draw, home win.

Thus we measure home advantage at event-level detail. Many studies, for example Nevill and Holder (1999), calculate an average at a sports-league-seasonal level: the percentage of matches won by the home team in all matches during a season. The implicit argument is that relative team quality is fixed throughout a season, and a balanced schedule exists where each team plays every other team once at home and once away. If this is not the case (for example, if teams can alter playing squads mid-season, if teams or players develop during a season, or if the schedule is not balanced), then it may be that such an aggregation provides a biased measure of home advantage $\alpha$ since it implies the restriction $\beta = 1$. Unbalanced schedules are very common in sports leagues; Lenten (2008b) considers their impact on conventional measures like competitive balance in the top division of Scottish football, and Lenten (2011) investigates unbalanced scheduled in Aussie Rules Football. English cricket’s County

2Though note that in cricket Duckworth and Lewis (1998) have devised a method to quantify expected outcomes in the case of rain delays, which can be adapted in order that runs/wickets results can be converted into a common set of units (Clarke and Allsopp 2001, see, e.g.), Jewell and Read (2014) employ a slightly different method for this conversion.

3In cricket, a team batting second and chasing a total could win by between one and ten wickets, but as Clarke and Allsopp (2001) argue, such margins do not express the true margin of victory since the chasing team stops batting once they have reached the target, and may have achieved a significantly larger number of runs.

4A tied outcome in cricket is not the same as a tied outcome in many other sports. A tie would represent both teams scoring an identical number of runs. As with high-scoring games like basketball and rugby, this is very rare an outcome. Much more common is a draw, where in the allotted time available (which may be reduced by adverse weather), neither team was able to bowl the other team out and claim victory.
Championship has been unbalanced for much of its existence.

In their subsequent analysis of golf and tennis tournaments, Holder and Nevill (1997) control for the strength of contestants in a competition in order to identify home advantage. If we define home advantage to be home teams winning more often than their ability suggests, then controlling for team strength is essential.

The $x_i$ variable of relative strength needs to be constructed. A common strategy to do this is to use betting odds since these contain information on the expected outcome of a match (see, for example, Croxson and Reade (2014) on the information efficiency of betting odds, but also Sobel and Ryan (2008) on the known biases of betting odds). However, bookmaker odds only exist so far back, historically, and only for matches in a small fraction of competitions that exist even in recent history, and hence in order to be able to measure relative quality more broadly we need an alternative measure of the strength of teams. As we consider teams across a wide range of countries and competitions, we employ Elo rankings.

Elo rankings provide a strength rating for each participant in a contest, and were developed for the purpose of ranking chess players. These strength ratings are then used to create predictions of future event outcomes.

Each team has a strength, $x_{i,H}$ for the home team, and $x_{i,A}$ for the away team. Conventionally a team has strength 1000 in its first match. For match $i$, a prediction of the outcome can be generated in terms of the expected score for each team, according to the formula:

$$x_i = \frac{1}{1 + 10^{(x_{i,A} - x_{i,H})/400}}.$$  

(3)

The formula for $x_i$ is symmetric in relative team strength, and $0 < x_i < 1$, where the extreme cases represent one infinitely strong team against an infinitely weak team. In practice, the largest difference between two teams in the cricket data is around 700. We use the outcome variable $y_i$ as in (2) to update the Elo strengths for each team by the formula:

$$x_{i+1,H} = x_{i,H} + K(y_i - x_i).$$

(4)

Here, $K$ is chosen to determine the sensitivity of the ratings to each individual match. A range of values have been used for this. We simply use $K = 40$.

There is no explicit role for the existence of home advantage in calculating either the strengths, or event predictions, but both will reflect home advantage implicitly, if participants win more often when playing at home. Assuming that teams play a roughly proportionate number of matches at home as away, however, their Elo strength should not reflect home advantage, and the Elo prediction should reflect the actual quality differences between teams.

Our interest is in the size of the home advantage, which is dictated by the $\alpha$ and $\beta$ coefficients, but in an ordered probit/logit model, the constant term is somewhat trickier to interpret. Given that ordinary least squares estimation of (1), also referred to as a linear probability model, will still yield unbiased estimates in the presence of a discrete dependent variable, and given that is it most intuitive to understand, we use this method in this paper. We present some results as equations, and in other

---

5 The slight drawback of this approach is that it provides a single ranking for cricket teams that play different forms of the game, while it may be that some teams specialise in one form (say, one-day) relative to another.

6 Naturally, a small number of matches is needed for each team in order for their strength to calibrate to something nearer to their actual strength.
places comment on the value of the $\alpha$ and $\beta$ coefficients, and also plot the implied regression lines.

3 Illustrating the Literature

In this section, we highlight the historic persistence of home advantage. We look at data for cricket back to the eighteenth century. In doing so we address a number of important aspects of the measurement of home advantage, covered in the literature.

We will illustrate using a vast dataset of cricket matches. The Cricket Archive website provides scorecards for cricket matches played throughout the world of cricket, from games at the very top of the game down to regional leagues in cricket playing nations. The website covers some of the earliest recorded matches in England in the early eighteenth century, although the richness of the data at regional level is more of a recent phenomena.

Most previous studies consider a single sports league over a period of time. Rather than restrict attention, we consider all recorded cricket matches, which covers 11,441 leagues and cups in 109 countries, as well as a huge number of ad hoc matches such as tour matches. There are mens and womens matches, youth, full, and also veterans matches. In total, Cricket Archive has 610,208 cricket matches that we can use to conduct our analysis of home advantage. Naturally, the number of matches has increased over the years; there is 379 recorded scorecards from the 18th century, 29,961 from the 19th century, 178,061 in the 20th century and 401,809 in the 21st century. Cricket before the 19th century was generally played on an ad hoc basis, but during the course of the 19th century test cricket between cricketing nations (nations of the British Empire) and domestic leagues began to form, and this development continued into the 20th century.

The first step is to establish the existence of home advantage. Previous studies of home advantage have considered a wide range of sports and sports leagues. Schwartz and Barsky (1977) consider baseball, basketball, American football and ice hockey in 1971, and Pollard and Pollard (2005b) add English football and a substantially longer set of years to their study. Schwartz and Barsky (1977) find home advantage to be strongest in basketball and hockey, and weakest in baseball and American football. Pollard and Pollard (2005b) show that home advantage has been declining across all sports except baseball and American football, where the latter shows a great degree of volatility. As such the distinctions observed by Schwartz and Barsky (1977) have largely disappeared, although baseball still does retain a relatively most (and consistent) advantage.

Gómez et al. (2011) investigated nine different team sports in Spain: baseball, basketball, handball, indoor soccer, roller hockey, rugby, soccer, volleyball, and water polo. They found a significant home advantage in all, although strongest in rugby.

In Figure 1 we plot the proportion of matches ending in wins for the home team and away team per year, since 1775. It identifies something Pollard and Pollard (2005b) note for many sports: in the earlier years of their existence, home advantage is stronger. In general, it is stronger in cricket, but it also shows much greater volatility. This reflects in part the fewer number of matches in early years, and also their irregularity; it was not until 1890 that the English County Championship began, for example.

The dashed line is the proportion of home wins, and hence if this is higher than the proportion of away wins (dotted line), this is indicative of home advantage. Such a graphical analysis is limited since

---

7See http://cricketarchive.com
8Just 31 recorded matches occur between 1731 and 1775, and hence given their sparcity over the years, we omit to plot these outcomes.
Figure 1: Relative proportion of matches ending in wins for the home or away team, or ending without a decisive result. Source: Cricket Archive.
the matches covered in any calendar year do not constitute a balanced schedule. Instead this is all recorded cricket matches in a calendar year. Nonetheless, a familiar pattern is visible: the dashed line is consistently above the dotted line with few exceptions.

Those exceptional periods appear to correspond to times when the proportion of matches ending without a decisive outcome is higher (for example in the late nineteenth century and the 1950s and 1960s).[9] Lenten (2008a) discusses the decline in the frequency of drawn outcomes since the 1960s, arguing that it increases the predictability of outcome, and hence may reduce interest in some forms of the game of cricket.

As Figure 1 does not reflect a balanced schedule, then as Holder and Nevill (1997) note, we need to control for team strength. Equation (3) provided a method of generating a forecast for each match outcome, and rather than summarising only the outcomes (using \( y_i \)). Hence in order to better reflect the relative strengths of teams, in Figure 2 the proportion of matches won by the home side when the home team is stronger, and the proportion of matches won by the away team when the away team is stronger are plotted. In the absence of home advantage, these two proportions should be similar. In the very early period, there is a lot of noise, but from around 1790 onwards, the series begin to settle down, with stronger home teams winning around 15% more games than stronger away teams, indicating a substantial home advantage. In around 1810 this disparity fell, as between then and around 1840, stronger home teams won only about 6% more than stronger away teams. After this, and through most of the twentieth century, there was no distinguishable difference between the two ratios, indicating an absence of home advantage once team strength is accounted for. This is consistent with Pollard and Pollard (2005a) who found that across a range of major team sports, home advantage was strongest in the earlier years of their existence. Since the late 1980s however, stronger home teams have won around 3% more than stronger away teams. This is in contrast to the home advantage pattern in English football that Koyama and Reade (2009) identify: in football, since the late 1980s home advantage has fallen. They attribute that to television, which enabled football spectators to better monitor the effort levels expended by players in away matches.

Quantifying that aggregate home advantage, we estimate (1) on all 610,208 cricket matches to yield (standard errors in parentheses):

\[
y_i = 0.136 + 0.784 x_i (0.002) (0.003)
\]  

(5)

This regression line is plotted in Figure 3, along with scatter points reflecting the proportion of matches for a given relative quality difference [10]. The slope coefficient is significant, and also significantly less than one, while the constant term is significantly different from zero, together indicating the presence of a home advantage. As the slope is flatter than the 45-degree line, this indicates that at some point a home team becomes less likely to win — when the regression line intersects the 45-degree line. Hence the home advantage in cricket favours weaker teams, for the most part. When the Elo prediction is above 64%, fewer home teams win that the quality difference would suggest should. In our dataset, 18.6% of matches fall into this category. Hence in 81.4% of matches over 286 years, the home team wins more often than it ought to given its relative strength: home advantage.

---

[9] The graph displays that these matches were drawn. In this category is all matches not resulting in a decisive outcome for one of the two teams competing. This could be due to all sorts of reasons, not least weather.

[10] We group quality differences to the second decimal place, and average outcomes in that range, and plot the resulting point. Hence the fitted line plotted is not fitted to the points that are plotted, but rathe all of the 0s, 0.5s and 1s representing actual outcomes. The points are plotted to get a sense of the linearity or otherwise of the underlying relationship between relative team strength and average outcomes.
Figure 2: Relative proportion of matches ending in wins for the favourite team (according to Elo ranking).
Cricket Match Outcomes by Relative Quality

Figure 3: Plot of home advantage regression result from all cricket matches. Matches, along the horizontal axis, are collected in intervals of one percentage point of quality difference, such that matches where $x_i < 0.01$ are summarised into the first point, then matches where $0.01 \leq x_i \leq 0.02$ are summarised in the second point, reading left to right on the plot.
Even though a team is listed first, this does not mean that that team is necessarily playing at home. Conventionally, Cup finals are played at a neutral venue, as are World Cups and other tournaments. In many sports leagues globally, teams share stadia also, with the Australian Football League being something of an extreme case with 10 teams sharing two stadia. Pollard (2006) note this in their worldwide study of home advantage in football; if two teams that share a stadium meet, some of the aspects of home advantage cease to exist, namely travel and familiarity. Nonetheless, the crowd composition will vary depending on who is the home team, as may other aspects of a match (which dressing rooms and other facilities can be used by the home or visiting team).

The Cricket Archive website lists the venue for every match, with a venue identifier that includes reference to the country the venue is in. There are 11,637 matches between two country teams in our dataset, a further 2,731 matches where the home team is not a national team but the visiting team is (usually a tour match), and 2,469 where the away team is not a national team but the home team is. Importantly for identifying home advantage, of those 14,368 matches with a national team listed first, 9,500 were played in that country, meaning that the remaining 4,868 were played on neutral territory.

Of those played on neutral territory, there is essentially no home advantage. Figure 4 plots the regression line (the solid line), and it intersects the red 45-degree line at 0.51. The slope is slightly flatter than the 45-degree line, but this merely indicates that weaker teams win more often than their relative strength would indicate since the quality measure is symmetric around 0.5. For matches where the team listed first is at home, the regression line significantly shifts up by 0.08 such that the intersection with the 45-degree line is at 0.79. This means that only for extremely unbalanced matches, constituting about 4–5% of all matches in our dataset, is a home team not more likely to win given its relative strength.

It also means that we can identify the intercept as the size of the home advantage, since the slope relates to the uncertainty of outcome and applies equally to home teams as away teams. Given this, we can say that home teams win 8% more often than away teams of identical relative quality. As such, we are able to identify a strong home advantage after controlling for team ability, and accounting for the fact that many matches are played on neutral territory.

This is, however, not particularly interesting without an explanation. Studies into home advantage have proposed explanations that can be classed under four broad headings, Nevill and Holder (1999) suggest: crowd, learning, travel and rules. Crowd-based mechanisms tend to focus on the size of the crowd in attendance and its impact on outcomes, focussing on two mechanisms: influencing player performance via encouragement, or influencing the actions of officials. Under learning, familiarity with local conditions (pitch size, stadium, climate/geographic factors) is important. The fatigue associated with travel, and the disruption to normal routines, is argued to influence performance and hence may well contribute to home advantage. Finally, it may be that rules influence outcomes in favour of home teams. In the next four subsections we explore these four explanations.

3.1 Crowd

Two aspects of crowds would appear to matter: it’s size, but also its density. A few thousand spectators at a tennis match or ice hockey game might create a more intimidating atmosphere for visiting players than a couple of thousand spread around a cricket or baseball field. Two mechanisms for the crowd influencing outcomes have been proposed; firstly a direct encouragement effect. Numerous papers, such as Schwartz and Barsky (1977), make the point that intuitively it is suspected crowd matters, since

\[\text{We also added a dummy for teams listed second being at home, which happens 188 times in our dataset. The results are essentially identical.}\]
Figure 4: Plot of home advantage regression result from international cricket matches. Matches, along the horizontal axis, are collected in intervals of one percentage point of quality difference, such that matches where $x_i < 0.01$ are summarised into the first point, then matches where $0.01 \leq x_i 0.02$. are summarised in the second point, reading left to right on the plot.
encouragements to attend and support teams are commonplace. [Wolfson et al. (2005)] show that sports fans believe their presence affects outcomes. A second, more quantifiable mechanism is that the crowd influences the decisions of officials that then influence match outcomes.

[Nevill et al. (1996)] note that red card and penalty decisions appear influenced by crowd size, although the direction of causation is not necessarily clear, since if home sides attack more, that pattern of play is consistent with more red cards for away teams, and more penalties for home teams. [Buraimo et al. (2010)] attempt to control for this by considering a minute-by-minute analysis of yellow and red cards in-match. They do find that home teams get fewer disciplinary sanctions, but in addition find that this effect is absent for teams (in Germany) who had stadia with running tracks, where the crowd would be much further from the field.

[Pettersson-Lidbom and Priks (2010)] consider a special case of football matches in Italy played without crowds. They find, consistent with the literature, that referees award more fouls and cards against visiting players when crowds are present. Cricket has its own unusual events that enable analysis of home advantage. Afghanistan have never played ‘home’ matches in Afghanistan, instead playing in the United Arab Emirates and India, and since 2011 Pakistan have played home matches in the United Arab Emirates. We can investigate these cases by including dummy variables for Afghanistan and Pakistan post-2011. Neither team appear to display any significantly different set out outcomes in matches from the average.

Cricket is not particularly renowned for large crowds, and given the size of the playing area, the density of crowds and proximity to the players is low. Nonetheless, big international matches attract considerably larger crowds than domestic matches. Mens matches attract considerably larger crowds than either youth or womens’ matches, and one-day matches and shorter matches relatively larger crowds than test matches (which are played over a number of days). In recent years domestic 20-over (T20) leagues have attracted relatively high attendances, not least because such games take place over a three-hour period. In particular, domestic leagues in India, Bangladesh and Australia have attracted average attendances of around 20,000. Indeed, there have been news articles discussing the kind of intimidation being faced by cricketers at domestic T20 matches. T20 matches can be grouped into three: domestic leagues and cups, tour matches and international tournaments (the World Cup). We consider all T20 matches, and insert dummy variables for domestic leagues, and in particular the three listed leagues with high attendances. When we do so, we find that home advantage increases dramatically. This is presented graphically in Figure 5 where the dashed line is the domestic matches and the solid line is international matches. This doesn’t necessarily shed light on the mechanism — a more detailed analysis of decisions such as LBW would be necessary to investigate the officials mechanism — but it provides evidence to suggest that the crowd does matter.

Historical data on attendances exists for Leicestershire and Lancashire County Cricket Clubs, but otherwise attendance data is relatively lacking in cricket. Regressing match outcomes on attendance for both counties produces an insignificant coefficient.

---

12 Wikipedia lists sports by attendance (see [https://goo.gl/9tbG6c](https://goo.gl/9tbG6c)), and while cricket does not figure as a sport in any of the lists of attendances, nonetheless two of the top ten sports leagues for average attendance in 2015–16 are cricket leagues: the Indian Premier League, and the Big Bash League in Australia.

13 See, for example, “Peter Trego: Somerset all-rounder says T20 abuse at ’football hooligan levels’", [BBC Sport, 18 August 2017](http://www.bbc.co.uk/sport/cricket/40974447).

14 [Reade (2017)] estimates a daily demand function for Leicestershire using some of this data.
Figure 5: Plot of home advantage regression result from all T20 cricket matches, grouped by whether a match is international, or a domestic league or cup match.
3.2 Familiarity

Familiarity with the venue at which a match is being played is argued to be a factor in home advantage. The dimensions of playing area, the facilities, the local area, may all play some part consciously or otherwise. Pollard and Pollard (2005a) note that after the Second World War home advantage was significantly lower in English and Italian football, when players would have been less familiar given a long break. Pollard (2005) found that teams moving to a new stadium in baseball, basketball and ice hockey in North America lost 24% of the advantage associated with playing at home initially.

In cricket, there is plenty of reason to suspect familiarity plays a role. Schwartz and Barsky (1977) lists a similar set of reasons for baseball: Ground staff may prepare a pitch that favours home players, and the size of the playing area ensures players must be familiar with its dimensions and character. All of these could be factors in why we observed a significant home advantage earlier. However, there is no significant postwar effect on home advantage, nor is there much experience with new stadia to compare with football. County teams in England do play at a number of venues, conventionally. Counties play the majority of their matches at their primary venue, but a proportion of matches at one or more out fields: smaller grounds elsewhere in the county. Attendance analyses (Schofield, 1983; Paton and Cooke, 2005) have suggested that such out fields do attract greater crowds, not least because they are often associated with festivals and are one-off annual events. Nonetheless, they represent alternative venues from the ones that home players should be most familiar with. Inserting a dummy which is one when a county plays at a venue other than its primary ground yields marginally significant results, suggesting that indeed, home advantage is slightly smaller where familiarity is slightly lower. But this lack of familiarity might be tempered by the larger crowds such matches attract.

Familiarity may vary with the length of matches. Cricket matches vary from the short to the very long; historically, even timeless matches existed (1,339 recorded in our dataset), but the most common match is a single-innings one-day match (more than half of our observations). There are 61,879 3-day matches, 41,201 2-day matches, and 20,234 4-day matches. Conventionally, test matches are five days in length, and there are 2,610 such length matches recorded.

With a multiple-day match, visiting teams get more opportunity to familiarise themselves with local conditions, and as such we might anticipate that home advantage is smaller in longer-form matches. Equally, of course, it may be that by the time familiarity has been acquired, the outcome of the game has already been determined. Shorter cricket matches are distinguished by being played over limited overs: a fixed limit on the number of balls to be bowled per innings, rather than an innings continuing until a team has been bowled out (or the batting team declaring). If we consider only limited overs matches, as in Figure 6, then we find that home advantage is reduced. The dashed line is limited overs matches, and the solid line is unlimited overs matches. The limited-overs line has a steeper slope, and a smaller vertical intercept, indicating a smaller home advantage. The two lines pivot around 0.5, meaning that at the upper end of the scale where the home team is much stronger, outcomes are closer to what quality differences imply they should be. Hence home advantage is stronger in longer matches, where visiting teams have more time to become familiarised with a venue. This suggests that familiarity may not be an important factor in determining home advantage; but it also may indicate that by the time familiarity has been acquired, the match is already too far developed for the outcome to be influenced.

In the last half century or so, the number of concurrent tournaments taking place in countries has led to some teams playing each other in different formats at very similar points of the season. For example, in England the advent of a one-day tournament on Sundays in 1969 added to an already busy

\footnote{In cricket, pitches are well known for deteriorating as play progresses.}
Figure 6: Plot of home advantage regression result from all cricket matches, grouped by whether a match is limited in overs or not. Outcomes are summarised by the relative quality difference of a match.
weekly calendar of three-day matches starting on Saturdays and Wednesdays (with a break on Sunday, traditionally). Hence it was often the case that two teams would begin a three-day match on a Saturday, then on a Sunday play a one-day match at the same venue, and on Monday resume the three-day match. If familiarity plays a role in home advantage, it should be that in such one-day matches the home advantage is reduced. More generally, we might anticipate that home advantage is increasing in the number of days between meetings between teams at a particular venue.

Calculating the days between meetings and adding this as an explanatory variable ($\Delta t$) and as an interaction term with relative team quality provides some evidence regarding this:

$$
y_i = 0.145 + 0.769 x_i - 17.255 \times 10^{-7} \Delta t + 30.214 \times 10^{-7} x_i \times \Delta t.
$$

(6)

Although both the intercept and slope effects are significant, the intercept dummy has the wrong sign: as the number of days increase between meetings the home advantage decreases (incredibly slowly).\footnote{It would take 10,000 days before a one percent effect was had on outcomes. The effect is even less significant if we restrict attention to just English County Championship matches where our motivating scheduling quirk occurred.}

Considering one-day games that begin on a Sunday between the same two teams that started on the Saturday, we find that for both the one-day and three-day games paired together like this, home advantage is significantly stronger than in standard matches. This would appear to argue against familiarity as a cause of home advantage, since in particular the one-day match begins the day after a visiting team has had ample opportunity to become familiar with the ground (certainly relative to if the one-day match did not begin the day after a three-day match had started).

Considering all aspects of familiarity covered in this section, the evidence appears slight that familiarity underlies home advantage.

Finally, it may be that familiarity is with weather and climatic conditions rather than physical surroundings. Anecdotal stories about of particular sporting venues with very particular weather conditions (lots of rain, very cold), but also about the local geography.\footnote{Bolivia is not one of the 109 countries in our dataset.} \cite{McSharry-2007} investigated the impact of playing football at high altitude in response to claims that Bolivia gained from an unfair advantage in South American football matches, finding that indeed altitude mattered.\footnote{We have 11,637 international matches, and 95,462 matches that we identify as regional.} Conversely, \cite{Chumacero-2009} find that altitude does not matter for determining home advantage, although heat and humidity do.

### 3.3 Travel

Travel induces fatigue which, via jet lag, can be long lasting. The travel hypothesis is that team performance is inhibited by travelling longer distances. If this is true, it would be expected that the home advantage is stronger for international matches where teams are often travelling across multiple time zones, and weaker for regional matches.\footnote{Lenten and Winchester (2010) note an asymmetry in home advantage in Aussie Rules Football whereby teams travelling east across time zones suffer more than teams travelling west.} Lenten and Winchester (2010) note an asymmetry in home advantage in Aussie Rules Football whereby teams travelling east across time zones suffer more than teams travelling west.

In Figure 7 we present evidence regarding this hypothesis by splitting up the matches into international, mixed and regional.\footnote{It would take 10,000 days before a one percent effect was had on outcomes. The effect is even less significant if we restrict attention to just English County Championship matches where our motivating scheduling quirk occurred.} The red line is the 45-degree line, which we would expect observations to lie around if there was no home advantage. The solid line is non-international, non-regional matches, and is similar to that plotted in Figure 3. The dashed line for international matches is slightly steeper, but
Figure 7: Plot of home advantage regression result from all cricket matches, grouped by travelling distances. Outcomes are summarised by the relative quality difference of a match.

The steady improvement in the quality of travel, and its affordability coincides with the steady decline in home advantage documented across multiple sports by Pollard and Pollard (2005b). Describes the travel arrangements on an England tour to India in 1951/52. The following text from is from Allan Watkins: “The journey to Karachi ahead of the second ‘Test’, as Allan recounts it, gave the players different reasons for concern: “We went in one of those old cargo planes. The pilot took us down and we went in to land at the ordinary airport, but they wouldn’t let us down. So we had to go to their RAF airfield. We went wobbling across Karachi and when we came down we finished up facing the wrong way. The plane went round on its wheels after landing. The poor old pilot wouldn’t come back – they usually came back and we thanked them, but he wouldn’t come back. We learnt later that it was the first time the poor bugger had flown at night.”

Equally, however, with much longer trips being necessary for tours (for example, the 1958/59 English tour to Australia set sail on September 27 and docked in Australia on October 13), some players were prevented from participating for financial reasons, and as such touring teams might...
not be as strong. If we surmise that the year 1965 was the point after which touring teams flew rather than sailing, and focus only on tour matches, we find evidence for a slight effect of air travel in favour of increasing home advantage. This would appear to argue that it is jet lag, rather than the travelling per se, that contributes towards home advantage. While the distance travelled varies considerably between these categories of matches, so equally does the absolute quality of the teams, the length/type of match, the decisiveness of the match (knock-out vs league), and more than likely, the crowd size. International matches draw the largest crowds, followed by national matches then regional ones. Within each type of match, shorter matches draw larger crowds, as the tendency towards shorter forms of the game in recent years betrays. Fixing the type of match and repeating the above analysis appears to make little difference — regional matches do appear to show greater home advantage than international matches.

Another aspect of note with travel is that home players may also need to travel to reach the home venue. Players who have recently moved football clubs often will not immediately relocate. In addition, in cricket county teams would draw players from throughout the county, and up until the 1960s car ownership was reasonably scarce. This also applied for national teams playing touring sides.

3.4 Rules

The purpose of rules are to make contests as fair as possible between the teams participating, and hence in principal should mitigate against home advantage. The purpose of officials is to ensure the rules are implemented such that contests are fair, hence the purpose of officials is also to mitigate against home advantage. Nonetheless it is possible that at times rules have failed to mitigate against home advantage, and also that officials have.

In cricket the decision on whether to bat or bowl first is determined by coin toss, with the home team much more knowledgeable should they win the toss than the visiting team. In 2016, the English County League began an experiment with uncontested tosses, where the visiting team can choose to bowl first, or have a standard coin toss. As this experiment is in its infancy, it is probably too early to determine whether it has had an impact on home advantage.

Another example of rules is permitting home teams discretion in important aspects of the hosting of an event, officials can also create further home advantage. For example, the selection of a venue at which to play may be determined based on factors that make a home win more likely. In tennis’s Davis Cup, for example, the host nation can choose both the venue and the surface that a match will take place on. Interestingly, Gayton et al. (2009) investigate the home disadvantage in Davis Cup matches, finding that based on Final matches between 1900 and 2006, in decisive fifth matches the away team

---

21 Additionally, injured players may be less likely to be selected for tours.
22 We note that touring sides could still choose to sail, but do not. An associated factor, which the website test-cricket-tours.co.uk details is that tours are much shorter now than they were. The 1962/3 England tour to Australia was 135 days in length, with 52 days between arrival in Australia and the first test. The 2006/07 tour was 102 days in length, and there was just 18 days between arrival (by plane) and the first test.
23 Though cars were used for many of the journeys, away matches often meant travelling by train, and it could be late in the evening before the players were back in Wales. “We’ve often finished a game in London and scrambled for a train at Paddington and stood in the corridor all the way down. We might get a seat at Cardiff. The trains were unbelievably full in those days, but we got to know the chap in the dining car and sometimes he’d put us down for the second sitting. And then you could sit there because he wasn’t serving any more meals. Stan was his name” Even if it was past midnight before the team’s train reached Cardiff, the county still offered no help with accommodation, so those who lived further afield might have to stay at a hotel. “There was a horrible dingy thing by the station, and noisy as well” Don remembers, “but we had to pay for our own digs if we were starting in Cardiff next day.”
Table 1: Proportion of cricket test matches won by each team.

<table>
<thead>
<tr>
<th>Test Number</th>
<th>No. observations</th>
<th>Away</th>
<th>Draw</th>
<th>Home</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>691</td>
<td>26.6</td>
<td>30.9</td>
<td>42.3</td>
</tr>
<tr>
<td>2</td>
<td>554</td>
<td>27.6</td>
<td>34.4</td>
<td>37.8</td>
</tr>
<tr>
<td>3</td>
<td>382</td>
<td>23.5</td>
<td>37.1</td>
<td>39.2</td>
</tr>
<tr>
<td>4</td>
<td>183</td>
<td>25.0</td>
<td>34.8</td>
<td>39.7</td>
</tr>
<tr>
<td>5</td>
<td>130</td>
<td>25.2</td>
<td>34.4</td>
<td>39.7</td>
</tr>
</tbody>
</table>

Table 1: Proportion of cricket test matches won by each team.

It is worth noting that there are only 21 such matches as all other Finals were concluded without requiring a fifth match. Nonetheless 13 of those 21 resulted in wins for the away team. The authors suggest that this is consistent with the idea that the pressure of the crowd inhibits players from carrying out skillful activities, and is something that [Baumeister and Steinhilber (1984)] also found in baseball’s World Series decisive playoff games. However, [Gayton et al. (2009)] do not control for the strength of home or away teams.

The nature of tests in cricket enable an investigation of this with our data. Test matches take place between qualified teams (there are only twelve of them, recently increased from ten with the addition of Afghanistan and Ireland), and like play-off matches, will be a succession of matches between the same two teams. For example, conventionally England play Australia over a series of five-matches. In our dataset, we have 1947 test matches, of which 691 are first tests, 554 are second tests in a series, 382 are third tests, 183 are fourth tests and 130 are fifth tests, reflecting that different series will be of different lengths. Table 3.4 shows the proportion of tests won by the home side, away side, and how many were drawn. On the basis of these proportions, home advantage appears to decline slightly, as 42% of first tests are won by the home side to just 26.6% of away sides, whereas 39.7% of fifth tests are won by the home side and 25.2% by the away side. However, this analysis does not factor in the strengths of teams involved, and judging by the falling number of games as we move from the first to second test, the strength of the teams competing will not be identical.

In Figure 8 we plot the results of running regression (5) for first test matches, second test matches through to fifth test matches. This controls for team strengths, and presents a more nuanced picture. The red line is the 45-degree line, which represents an absence of home advantage. The home advantage in first test matches (solid line) is consistent throughout the range of relative quality levels between the teams, and suggests that home teams in tests win about 9% more than their relative ability suggests they should. For second tests, the slope flattens such that for test matches with a stronger home team, the home advantage lessens. For third, fourth and fifth tests, the home advantage shows a different relationship: weaker home teams gain a significantly greater advantage, as the regression lines are above the 45-degree line until the relative quality gets to around 0.65, and after that, for particularly strong home teams, there is a home disadvantage. Hence this lends some support to the home disadvantage theory set out by [Baumeister and Steinhilber (1984)] and [Gayton et al. (2009)], as stronger teams perform worse at home.

One feature of Davis Cup tennis is that although national teams face each other, the event consists

---

24 Davis Cup ties take place over up to five matches, or rubbers. The team to reach three wins first wins the overall event. The first two matches are singles matches, the third is a doubles match and the last two are singles matches.

25 An alternative explanation is that particularly by the fourth and fifth tests, the result may be beyond doubt, if a stronger team has won the first three tests, and as such the stronger team may try less hard, or experiment in the final two matches.
of up to four singles matches, and one doubles match. Hence it represents a mix of team and individual sports. The majority of investigations of home advantage focus on team sports such as football, baseball and basketball. A small number have focussed on on individual sports; Koning (2011) looks at tennis and finds that home advantage is strongest when two highly ranked players face each other, and disappears completely when two weak players face each other. Holder and Nevill (1997) consider tennis and golf, finding little evidence of home advantage.

Officials in charge of matches are expected to show neutrality in their application of rules. Convention dictates that officials should thus not hail from the region that a team is based. Although this does not guarantee an official will be unbiased, it may make it more likely: Gallo et al. (2013) show that referees in football implicitly discriminate. That is, subconsciously they make decisions that discriminate against particular types of players. In addition, Dawson and Dobson (2010) document that referees in football of different nationalities display distinctly different patterns in their implementation of the rules in international football.

In many international matches in cricket, officials are drawn from the home country. Before 1992, all officials would be from the home country, but after 1992 one of the two umpires in test cricket matches was drawn from a third country (on a trial basis, with it being confirmed in 1994), and after 2002 both were from a third country. In one-day international (ODI) matches, until 2001 all officials were drawn from the home country, and since then one of the two umpires has been drawn from a third country. In T20 international matches, all officials are again drawn from the home country.

The possibility that this institutional arrangement contributed to home advantage can be investigated using the framework we have set up. We restrict attention to only test matches, and include dummy
variables for 1992, 1994 and 2002, and allow these dummies to affect the slope as well as the intercept. We find only marginal evidence of significance, albeit with coefficient signs in the anticipated directions (home advantage falling). We also investigate the impact of the 2002 change in ODI matches and again find coefficient signs consistent with declining home advantage with the change, but no significance. Ringrose (2006) and Sacheti et al. (2015) consider the test change in more detail. They both consider leg before wicket (LBW), a method of a player being bowled out that is open to greater levels of discretion, as it historically involved the umpire determining a counter-factual: what would have happened to the ball had it not hit the batsman’s leg? Both studies found that dismissals by LBW fell for the away side around these two changes (partially falling after 1994, then falling to no difference between the teams after 2004). Hence the impact on individual decisions can be clearly shown, whereas the impact on overall outcomes of matches is more muted. This suggests that the observed home advantage was not determined solely by this particular aspect of bias towards a home side. It would appear that other factors mitigated this bias.

Garicano et al. (2005) find that in Spanish football, referees systematically add more injury time to matches in which the home team is losing, and less injury time when the home team is winning. Pettersson-Lidbom and Priks (2010), Nevill et al. (1996), Sutter and Kocher (2004) and Boyko et al. (2007) all find evidence that refereeing decisions contribute to home advantage. Along with injury time decisions, incidence of penalties and cautions appear to favour the home side also. Nevill et al. (2002) showed qualified referees videos of incidents in football matches, and found that when the sound was turned on, the referees called 15.5% fewer fouls against the home team compared to when watching the videos in silence. There is thus a considerable body of evidence to suggest that officials are unable to act in a neutral manner, whether explicitly or otherwise. As such, even neutral officials in international matches will likely influence outcomes subconsciously. A consistent pattern in home advantage across sports is its decline through time, and hence if the actions of officials are responsible for home advantage, this requires that via some mechanism the impact of these actions is being gradually mitigated over time. In many sports in recent years video technology has been implemented to assist officials, yet the decline in home advantage precedes any such adoption of technology.

Figure 1 provides an interesting insight into the role of rules. The increase in draws in the nineteenth century corresponds with a decrease in home advantage, while the decrease in the frequency of draws since the 1960s has also occurred in line with an increase in home advantage. Rule changes that provoke more attacking play and thus reduce the likelihood of a draw have been cited as reasons for declining home advantage in football (e.g. additional substitutions, and three-points for a win, see Jacklin (2005)). However, as Koyama and Reade (2009) note, incentives for more attacking play affect both the home and away side equally and hence ought not to be reasons for a change in home advantage.

### 3.5 Other Explanations

Having considered the crowd, familiarity, travel and rules, we now consider a small number of studies with explanations that do not fit within these four categories. It is worth noting also the the home disadvantage noted by Baumeister and Steinhalber (1984) and Gayton et al. (2009) probably better fits under the psychological, rather than a rules-based mechanism for affecting outcomes.

Another more psychological explanation is territoriality; the perceived invasion of one’s territory. Pollard (2006) study home advantage across a huge range of countries, and determine that as home advantage is stronger in more mountainous regions (the Balkans and Andes), territoriality is a likely explanation for the phenomenon. A small number of cricket matches take place in these two mountainous
areas (117 matches), and there is no statistical difference between them and the rest of the matches. If the Middle East might be described as war torn, and hence likely to exhibit higher degrees of territorialism by the same argument [Pollard (2006)] employ, we actually find that of the more substantial number of cricket matches in this region (21,856), home advantage is significantly smaller. As such, there does not appear to be evidence for territoriality as an explanation for home advantage in cricket.

[Neave and Wolfson (2003)] consider a more biological explanation. They examine the role that hormones, and testosterone in particular, play in sporting outcomes. They measure testosterone levels in teams prior to matches, finding that home teams have higher testosterone levels than visiting teams. This raises the notion that home advantage may be stronger in mens contests relative to womens’ contests. It may also contribute to differences between youth and adult matches. In the cricket dataset we have 27,197 womens’ cricket matches, 5,244 youth matches, of which 2,250 are womens matches. When inserting a dummy variable for womens matches, it is significant, and eradicates home advantage entirely. When doing the same, but restricting to youth matches, there is no significant difference between male and female match outcomes. In youth matches the home advantage is less marked than in adult matches. Intriguingly, suggesting some kind of life-cycle explanation of home advantage, since 2014 an over-50s, over-60s and over-70s County Championship has run, and in the over-50s and over-60s the home advantage is very small, while in the over-70s competition there is no home advantage. These results are summarised in Figure [9] where a distinctly sexist and ageist colour scheme has been employed. Blue lines relate to mens results, pink to womens, and grey to veterans. The line furthest from the 45-degree line is the mens cricket regression line, with the other lines much closer.

Again, it seems likely that other important factors vary when considering youth, womens and mens, and seniors cricket, such as crowd sizes and commercial pressures. Nonetheless there appears scope for further research into biological explanations for home advantage.

4 Conclusions

In this paper the literature on home advantage has been surveyed, and new evidence from the sport of cricket has been presented to illustrate the findings of the literature. Home advantage exists and varies across sports, and in general appears to be in (slow) decline. As with the literature, our analysis using a wealth of cricketing data provide nothing conclusive in terms of a single factor explaining home advantage. There is some evidence for familiarity, there is evidence for rules and the role of officials, and potentially evidence surrounding crowd, psychological and biological factors.

References


26This effect is surprisingly robust to restricting the sample to more recent years when the womens game has expanded significantly.
Cricket Match Outcomes by Age and Sex

Figure 9: Plot of home advantage regression result from all cricket matches, grouped by whether a match is limited in overs or not. Outcomes are summarised by the relative quality difference of a match.


